HEPATITIS C VIRUS HELICASE CRYSTALS AND MOLECULES COMPRISING HELICASE BINDING POCKETS

TECHNICAL FIELD OF THE INVENTION

The invention relates to the X-ray crystal structure of the hepatitis C virus helicase domain. More specifically, the invention relates to crystallized complexes of HCV helicase and an oligonucleotide, to crystallizable compositions of HCV helicase and an oligonucleotide and to methods of crystallizing an HCV helicase-oligonucleotide complex. The invention further relates to a computer programmed with the structure coordinates of the HCV helicase oligonucleotide binding pocket or the HCV helicase nucleotide triphosphate pocket wherein said computer is capable of displaying a three-dimensional representation of that binding pocket.

BACKGROUND OF THE INVENTION

Infection by the hepatitis C virus (HCV) is responsible for most transfusion-associated cases of non-A, non-B hepatitis and also accounts for a significant proportion of community-acquired hepatitis cases worldwide. Relatively few HCV infected individuals

experience acute hepatitis, but up to 85% appear to develop persistent infection that often leads to chronic hepatitis and liver cirrhosis, eventually predisposing them to hepatocellular carcinoma. At present, vaccines are unavailable and no broadly effective therapies exist for this viral disease. Consequently, much research has focused on the HCV replicative enzymes as targets for more effective therapies.

HCV contains an approximately 9.6 kb single-stranded positive sense RNA genome classified as its own genus in the *Flaviviridae* family of animal viruses, which also includes the flavivirus and pestivirus genera. Its genome consists of a conserved 5' nontranslated sequence that serves as an internal ribosome entry site, a single open reading frame that encodes a polyprotein of >3000 amino acids, and a 3' nontranslated region. The 3' nontranslated region contains tracts of poly(U)$_n$ and poly(UC)$_n$ followed by a novel conserved 98 nucleotide sequence.

Proteolytic processing of the HCV polyprotein by virally-encoded proteases generates several nonstructural (NS) proteins with enzymatic activities essential for the replicative cycle of the virus [P. Neddermann et al., Biol. Chem., 378, pp. 469-476 (1997)].

NS2 encodes a presumed metalloprotease, NS5B is a RNA-dependent RNA polymerase, and NS3 is a bifunctional enzyme with a serine protease localized to the N-terminal 181 residues of the protein and a RNA helicase in the C-terminal 465 amino acids. The NS3 protease performs an intramolecular cleavage at the NS3/NS4A junction to form a tight noncovalent NS3-NS4A complex necessary for efficient processing of the remaining polyprotein [C. Failla et al., J. Virol., 69, pp. 1769-1777 (1995); R. Bartenschlager et al., J. Virol., 69, pp. 7519-7528 (1995); Y. Tanji et al., J. Virol., 69, pp. 1575-1581 (1995)]. To date, no evidence exists to suggest that the serine protease and helicase domains are separated by proteolytic processing of NS3 in vivo. This may reflect economical packaging of these enzymatic components, or could imply a functional interdependence between the two domains.

Numerous studies have demonstrated that the serine protease [J. L. Kim et al., Cell, 87, pp. 343-355 (1996); W. Markland et al., J. Gen. Virol., 78, pp. 39-43(1997).; C. Steinkuhler et al., J. Virol., 70, pp. 6694-6700 (1996)] and RNA helicase domains [J. A. Suzich et al., J. Virol., 67, pp. 6152-6158 (1993); C. L. Tai et al., J. Virol., 70, pp. 8477-8484 (1996); L. Jin et al.,

Arch. Biochem. Biophys., 323, pp. 47-53 (1995); and F. Preugschat et al., J. Biol. Chem., 271, pp. 24449-24457 (1996)] of NS3 can be expressed independently and isolated as catalytically active species. However, emerging evidence suggests that the NS3 protease and helicase domains may contact one another and modulate NS3 catalytic activities. Examples include apparent differences in pH optima of ATPase and RNA unwinding activities between a contiguous NS3 protein complexed with the NS4A cofactor [K. A .Morgenstern et al., J. Virol., 71, pp. 3767-3775 (1997); Z. Hong et al., J. Virol., 70, pp. 4261-4268 (1996)] and an isolated NS3 helicase domain [C. L. Tai et al., J. Virol., (1996), supra; L. Jin et al., Arch. Biochem. Biophys., (1995), supra; F. Preugschat et al., J. Biol. Chem., (1996), supra; and Y. Gwack et al., Biochem. Biophys. Res. Commun., 225, pp. 654-659 (1996)]. Similarly, the ATPase activities of both proteins differ in their sensitivity to polynucleotide stimulation. Contiguous NS3 appears to have a lower apparent dissociation constant for poly(U) than does the helicase domain [J. A. Suzich et al., J. Virol., (1993), supra; F. Preugschat et al., J. Biol. Chem., (1996), supra; K. A .Morgenstern et al., J. Virol., (1997), supra; A. Kanai et al., FEBS Lett., 376,

pp. 221-224 (1995)]. Aside from these differences, both proteins display nearly indistinguishable kinetic parameters for NTP hydrolysis when stimulated with saturating polynucleotide [J. A. Suzich et al., J. Virol., (1993), supra; K. A .Morgenstern et al., J. Virol., (1997), supra], both display 3′- 5′ directionality for translocation along a polynucleotide substrate, and the helicases of both proteins effectively unwind duplex RNA:RNA substrates [C. L. Tai et al., J. Virol., (1996), supra; Z. Hong et al., J. Virol., (1996), supra].

In addition to HCV, all flavi- and pestiviruses sequenced to date contain conserved helicase sequence motifs in their homologous NS3 proteins, suggesting that this enzyme plays an important role in the HCV replicative cycle [R. H. Miller et al., Proc. Natl. Acad. Sci. USA, 87, pp. 2057-2061 (1990)]. Consistent with this possibility, helicase encoding sequences have been identified in other viruses and helicases are suggested to catalyze the separation of double-stranded nucleic acid structures during transcription and genome replication [G. Kadare et al., J. Virol., 71, pp. 2583-2590 (1997)]. Previous studies with poliovirus, a positive-stranded RNA virus of the *Picornaviridae* family,

show that mutation of conserved sequence motifs in the 2C helicase inhibits virus replication and proliferation [C. Mirzayan et al., Virology, 189, pp. 547-555 (1992)]. Similar mutational studies on the helicases encoded by herpes simplex virus type 1 and bovine papilloma virus also show that these enzymes are critical for virus replication [P. MacPherson et al., Virology, 204, pp. 403-408 (1994); R. Martinez et al., J. Virol., 66, pp. 6735-6746 (1992)]. Thus, the ability to inhibit helicase activity in HCV may provide an avenue for the therapeutic treatment of HCV infection.

Unfortunately, little is known about the details of how ATP binding and hydrolysis leads to DNA or RNA strand unwinding by the helicase. Two structures of helicases crystallized in the absence of polynucleotide, but, unfortunately, they have not yielded the critical information needed to extrapolate to an enzyme mechanism [N. Yao et al., Nat. Struct. Biol., 4, pp. 463-467 (1997); H. S. Subramanya et al., Nature, 384, pp. 379-383 (1996)].

Thus, there is a great need to solve the crystal structure of the helicase complexed with an oligonucleotide and, in particular, to delineate the oligonucleotide and nucleotide triphosphate (NTP) binding

pockets of that enzyme. With this information, computer models of these binding sites can be created and potential inhibitors of HCV helicase can be rationally designed.

SUMMARY OF THE INVENTION

Applicants have solved this problem by providing a crystallized complex of the NS3 helicase domain of HCV and a single stranded oligonucleotide. That crystal has been solved by X-ray crystallography to a resolution of 2.2Å. Based upon that crystal structure, applicants have identified the key amino acid residues that make up the oligonucleotide binding pocket of the helicase.

Thus, the invention relates to a crystallized complex comprising the NS3 helicase domain of hepatitis C virus or mutants thereof and an oligonucleotide.

The invention also relates to crystallizable compositions comprising the NS3 helicase domain, either as an isolated polypeptide or as part of the full length NS3 HCV protein, or single amino acid mutants thereof, and an oligonucleotide. And it relates to methods of using such compositions to produce the aforementioned crystals.

The invention also provides the X-ray structure coordinates of an NS3 helicase-oligonucleotide complex. Those coordinates, or at least the portion that define the oligonucleotide binding pocket or the NTP binding pocket are useful in methods for designing inhibitors of the NS3 helicase, which in turn may be useful in treating HCV infection. This is another aspect of the present invention.

In a related aspect, the invention provides a computer programmed with the coordinates of the NS3 helicase oligonucleotide binding pocket or the NTP binding pocket and with a program capable of converting those coordinates into a three-dimensional representation of the binding pocket on a display connected to the computer.

Finally, the invention provides a computer which, when programmed with at least a portion of the structural coordinates of HCV NS3 helicase and an X-ray diffraction data set of a different molecule or molecular complex, performs a Fourier transform of these structural coordinates of the helicase coordinates and then processes the X-ray diffraction data into structure coordinates of the different molecule or molecular complex via the process of molecular replacement.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 lists the atomic structure coordinates for NS3 helicase in complex with dU_8 as derived by X-ray diffraction from a crystal of that complex. The following abbreviations are used in Figure 1:

"Atom type" refers to the element whose coordinates are measured. The first letter in the column defines the element.

"X, Y, Z" crystallographically define the atomic position of the element measured.

"B" is a thermal factor that measures movement of the atom around its atomic center.

"Occ" is an occupancy factor that refers to the fraction of the molecules in which each atom occupies the position specified by the coordinates. A value of "1" indicates that each atom has the same conformation, i.e., the same position, in all molecules of the crystal.

Figure 2 shows a diagram of a computer used to generate a three-dimensional graphical representation of a molecule or molecular complex according to this invention.

Figure 3 shows a cross section of a magnetic storage medium.

Figure 4 shows a cross section of a optically-readable data storage medium.

Figure 5 depicts a stereo ribbon diagram of the overall fold of the NS3 helicase with bound d(U)8. Domain 1 ~~is colored blue, domain 2 red, and domain 3 green. The sulfate and DNA are colored yellow.~~

Figure 6 depicts a superposition of domains 1 ~~(blue)~~ and 2 ~~(red)~~ of NS3 helicase based on conserved secondary structure motifs with a 2.0 Å C-alpha RMS deviation over 78 residues. Residues corresponding to binding to the 3' end of the oligonucleotide are depicted as thick lines. Also shown is the location of Trp-501.

Figure 7, panel A, depicts the residues surrounding the bound sulfate superimposed on the phosphate binding loops of eight deoxynucleoside monophosphate kinases. Panel B depicts the electron density encompassing the bound DNA substrate. The ~~orange color depicts~~ the difference Fourier (F_O-F_C) electron density map calculated before DNA or water molecules were built into the model. The ~~blue color~~ depicts the final $2F_O$-F_C electron density map calculated at 2.2Å resolution using the refined model.

Figure 8 depicts the secondary structure of HCV NS3 helicase (indicated above the sequence). The

conserved sequence motifs are underscored. The non-HCV N-terminal and C-terminal residues which were added during cloning are depicted in lower case. No density was observed for residues in italics. The residues are numbered based on their location in the NS3 protein.

Figure 9, panel A, depicts aligned sequences of conserved motifs from other helicases. The motifs are colored similar to what was previously reported for the PcrA helicase to aid in comparison of the structures of enzymes from superfamilies I and II. Panel B depicts the location of conserved motifs.

Figure 10 depicts a view into the central binding cleft of the NS3 helicase domain.

Figure 11, Panel A, depicts the effect of enzyme concentration and incubation time on HCV NS3 helicase. Panel B depicts the effect of incubation temperature on helicase activity. Panel C depicts the effect of pH on helicase activity. Panel D depicts the effect of monovalent cations on helicase activity. Panel E depicts the effect of ATP on helicase activity. Panel F depicts the effect of divalent cations on enzyme activity.

Figure 12, Panel A depicts the effect of enzyme concentration on the binding of [^{32}P]-ssRNA substrate to

HCV NS3 helicase. Panel B depicts the dissociation of pre-formed NS3 helicase/ [^{32}P]-labeled ssRNA complex by [^{3}H] -labeled ssRNA over time. Panel C depicts the effect of pH on the binding of ssRNA to helicase. Panel D depicts the effect of monovalent cation on ssRNA binding to helicase. Panel E depicts the effect of divalent cations on ssRNA binding to helicase.

Figure 13 Panel A depicts the effect of poly (U) on the ATPase activity of HCV NS3 helicase. Panel B depicts the effect of enzyme concentration on ATPase activity in the presence or absence of poly (U).

DETAILED DESRIPTION OF THE INVENTION

In order that the invention described herein may be more fully understood, the following detailed description is set forth.

According to one embodiment, the invention provides a crystallizable composition comprising NS3 helicase and an oligonucleotide.

The NS3 helicase protein in the crystallizable complexes of this invention is selected from the isolated helicase domain from any strain or the consensus sequence of the HCV NS3 protein (e.g., amino acids 167-631 of SEQ

ID NO:2); the entire NS3 protein from any strain of HCV or the consensus sequence of that protein (e.g., SEQ ID NO:2); any portion of the NS3 protein that contains a functional helicase domain, which has been indicated to be amino acids 183-582 by C. Hyun-Soo et al., J. Biol. Chem., 273, pp. 15045-15052 (1998), from any strain of HCV or the consensus sequence of that protein (e.g., amino acids 183-582 of SEQ ID NO:2, amino acids 167-631 of SEQ ID NO:2, amino acids 183-631 of SEQ ID NO:2) and amino acid mutants of any of the above (including, but not limited to, SEQ ID NO:2 or any portion thereof that includes amino acids 183-582 of SEQ ID NO:2 and contains one or more of the following single amino acid replacements: Ser231-to-Ala, Thr269-to-Ala, Ser370-to-Ala, Thr411-to-Ala, Trp501-to-Phe, Trp501-to-Leu or Trp501-to-Ala, Gln460-to-Ala, Arg461-to-Ala, Arg462-to-Ala, Arg464-to-Ala, or Arg467-to-Ala).

The NS3 protein utilized in the crystallizable compositions of this invention may also contain additional amino acids at the N- and/or C-terminus which may be useful in purifying the protein when produced recombinantly. For example, we have found that a poly-histidine tag at the C-terminus is useful in purifying NS3 proteins produced in recombinant host cells through

the use of appropriate resins, such as Q-Sepharose (Pharmacia, Piscataway, NJ). Such tags may also be useful in increasing the solubility of the NS3 protein.

The second component in these compositions is an oligonucleotide. Preferably, the oligonucleotide is single-stranded, although double-stranded oligonucleotides may be used and subsequently dissociated prior to crystallization. Preferably, the oligonucleotide is a polynucleotide of between about 6 and 20 bases. More preferably, the oligonucleotide is between about 6 and 12 bases. Most preferably, the oligonucleotide is polyuracil 8 nucleotides long (dU_8).

The molar ratio of NS3 helicase to oligonucleotide should be about 1:1, although ranges between 1:5 and 5:1 are acceptable.

The buffers and other reagents present in the crystallizable compositions of this application may be any components that promote crystallization and/or are compatible with crystallization conditions. An example of such a buffer condition is 15 mM MES (pH 6.5), 2.5 mM ß-mercaptoethanol.

The invention also relates to crystals of NS3 helicase complexed with an oligonucleotide. Both the NS3 helicase component and the oligonucleotide component are

the same as those described above for crystallizable compositions. These crystals are obtained from the above described compositions by standard crystallization protocols, such as the protocol exemplified in the Example section below.

The invention also relates to a method of making NS3 helicase-containing crystals. Such methods comprise the steps of:

a) obtaining a crystallizable composition comprising an NS3 helicase protein and an oligonucleotide in a molar ratio of between 1:5 and 5:1; and

b) subjecting said composition to conditions which promote crystallization.

Again, the choice for the NS3 helicase protein and the oligonucleotide utilized in the above crystallization method are the same as those set forth above for crystallizable compositions.

As mentioned above, applicants have solved the three-dimensional X-ray crystal structure of an NS3 helicase-dU_8 complex. The atomic coordinate data is presented in Figure 1.

In order to use the structure coordinates generated for the NS3 helicase-dU_8 complex or its NTP or oligonucleotide binding pockets or portions or homologues

thereof, it is often times necessary to convert them into a three-dimensional shape. This is achieved through the use of commercially available software that is capable of generating three-dimensional graphical representations of molecules or portions thereof from a set of structure coordinates.

Binding pockets, also referred to as binding sites in the present invention, are of significant utility in fields such as drug discovery. The association of natural ligands or substrates with the binding pockets of their corresponding receptors or enzymes is the basis of many biological mechanisms of action. Similarly, many drugs exert their biological effects through association with the binding pockets of receptors and enzymes. Such associations may occur with all or any parts of the binding pocket. An understanding of such associations will help lead to the design of drugs having more favorable associations with their target receptor or enzyme, and thus, improved biological effects. Therefore, this information is valuable in designing potential inhibitors of the binding sites of biologically important targets.

The term "binding pocket", as used herein, refers to a region of a molecule or molecular complex,

that, as a result of its shape, favorably associates with another chemical entity or compound.

Applicants have identified three binding pockets which are good targets for designing inhibitors. Two of these binding pockets reside in the region of the helicase where an oligonucleotide binds. These pockets are designated U4 and U8, based upon the nucleotide of dU_8 that lies in this pocket in an NS3 helicase-dU_8 complex. The third binding pocket is the NTP binding pocket. While this binding pocket has been partially described by others [T. Yao et al., Nat. Struct. Biol., 4, pp. 463-467 (1997)], applicants have further defined this pocket in a way that was not derivable from what was known in the art.

The terms "U4-, U8- and NTP-like binding pocket", as used herein, refer to a portion of a molecule or molecular complex whose shape is sufficiently similar to the NS3 helicase U4, U8 and NTP binding pocket, so as to bind common ligands. These commonalties of shape are defined by a root mean square deviation from the structure coordinates of the backbone atoms of the amino acids that make up these binding pockets in the NS3 helicase structure (as set forth in Figure 1) of not more

than 1.5 Å. The method of performing this calculation is described below.

In resolving the crystal structure of NS3 helicase in complex with an oligonucleotide, applicants have determined that NS3 amino acids Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501 and Tyr502 form close contacts (<4Å) with U8 of dU_8 in the NS3 helicase-dU_8 complex. Thus, a binding pocket defined by the structural coordinates of those amino acids, as set forth in Figure 1; or a binding pocket whose root mean square deviation from the structure coordinates of the backbone atoms of those amino acids of not more than 1.5 Å is considered a U8-like binding pocket of this invention.

Applicants have also determined that in addition to the NS3 amino acids set forth above, Pro230, Val256, Thr298, Ala497, Lys551, Gln552, Gly554, Glu555, Asn556 and Pro558 are within 8Å of U8 of the bound oligonucleotide and therefore are also close enough to interact with that substrate. Thus, in a preferred embodiment, a binding pocket defined by the structural coordinates of the amino acids that are within 8Å of U8 of the bound oligonucleotide, as set forth in Figure 1; or a binding pocket whose root mean square deviation from the structure coordinates of the backbone atoms of those

amino acids is not more than 1.5 Å is considered a preferred NS3 helicase U8-like binding pocket of this invention.

Applicants have further determined that the NS3 helicase amino acids that define the shape of the U4 oligonucleotide binding pocket are: His369, Ser370, Lys371, Tyr392, Arg393, Thr411, Asp412, Ala413, Cys431, Val432, Gln434, Ile446, Thr448, Arg461, Glu493, Glu555, Asn556 and Phe557. Thus, a binding pocket defined by the structural coordinates of these amino acids, as set forth in Figure 1; or a binding pocket whose root mean square deviation from the structure coordinates of the backbone atoms of these amino acids is not more than 1.5 Å is considered a NS3 helicase U4-like binding pocket of this invention.

Applicants have also more completely determined, as compared to the prior art, the NS3 helicase amino acids that define the shape of the NTP binding pocket. Those amino acids are: Pro205, Thr206, Gly207, Ser208, Gly209, Lys210, Ser211, Thr212, Lys213, Asn229, Ala234, Gly237, Phe238, Tyr241, Asp290, Glu291, His293, Thr322, Ala323, Thr324, Gln460, Gly463, Arg464 and Arg467. Thus, a binding pocket defined by the structural coordinates of these amino acids, as set forth

in Figure 1; or a binding pocket whose root mean square deviation from the structure coordinates of the backbone atoms of these amino acids is not more than 1.5 Å is considered a NS3 helicase NTP-like binding pocket of this invention.

It will be readily apparent to those of skill in the art that the numbering of amino acids in other isoforms of NS3 may be different than that set forth for herein. Corresponding amino acids in other isoforms of NS3 are easily identified by visual inspection of the amino acid sequences or by using commercially available homology software programs.

Each of those amino acids of NS3 helicase is defined by a set of structure coordinates set forth in Figure 1. The term "structure coordinates" refers to Cartesian coordinates derived from mathematical equations related to the patterns obtained on diffraction of a monochromatic beam of X-rays by the atoms (scattering centers) of a protein or protein-ligand complex in crystal form. The diffraction data are used to calculate an electron density map of the repeating unit of the crystal. The electron density maps are then used to establish the positions of the individual atoms of the enzyme or enzyme complex.

Those of skill in the art understand that a set of structure coordinates for an enzyme or an enzyme-complex or a portion thereof, is a relative set of points that define a shape in three dimensions. Thus, it is possible that an entirely different set of coordinates could define a similar or identical shape. Moreover, slight variations in the individual coordinates will have little effect on overall shape. In terms of binding pockets, these variations would not be expected to significantly alter the nature of ligands that could associate with those pockets.

The term "associating with" refers to a condition of proximity between a chemical entity or compound, or portions thereof, and a binding pocket or binding site on a protein. The association may be non-covalent -- wherein the juxtaposition is energetically favored by hydrogen bonding or van der Waals or electrostatic interactions -- or it may be covalent.

The variations in coordinates discussed above may be generated because of mathematical manipulations of the NS3 helicase-oligonucleotide complex structure coordinates. For example, the structure coordinates set forth in Figure 1 could be manipulated by crystallographic permutations of the structure

coordinates, fractionalization of the structure coordinates, integer additions or subtractions to sets of the structure coordinates, inversion of the structure coordinates or any combination of the above.

Alternatively, modifications in the crystal structure due to mutations, additions, substitutions, and/or deletions of amino acids, or other changes in any of the components that make up the crystal could also account for variations in structure coordinates. If such variations are within an acceptable standard error as compared to the original coordinates, the resulting three-dimensional shape is considered to be the same. Thus, for example, a ligand that bound to the oligonucleotide binding pocket of NS3 helicase would also be expected to bind to another binding pocket whose structure coordinates defined a shape that fell within the acceptable error.

Various computational analyses are therefore necessary to determine whether a molecule or the binding pocket portion thereof is sufficiently similar to the NS3 helicase binding pockets described above. Such analyses may be carried out in well known software applications, such as the Molecular Similarity application of QUANTA

(Molecular Simulations Inc., San Diego, CA) version 4.1, and as described in the accompanying User's Guide.

The Molecular Similarity application permits comparisons between different structures, different conformations of the same structure, and different parts of the same structure. The procedure used in Molecular Similarity to compare structures is divided into four steps: 1) load the structures to be compared; 2) define the atom equivalences in these structures; 3) perform a fitting operation; and 4) analyze the results.

Each structure is identified by a name. One structure is identified as the target (i.e., the fixed structure); all remaining structures are working structures (i.e., moving structures). Since atom equivalency within QUANTA is defined by user input, for the purpose of this invention we will define equivalent atoms as protein backbone atoms (N, Cα, C and O) for all conserved residues between the two structures being compared. We will also consider only rigid fitting operations.

When a rigid fitting method is used, the working structure is translated and rotated to obtain an optimum fit with the target structure. The fitting operation uses an algorithm that computes the optimum

translation and rotation to be applied to the moving structure, such that the root mean square difference of the fit over the specified pairs of equivalent atom is an absolute minimum. This number, given in angstroms, is reported by QUANTA.

For the purpose of this invention, any molecule or molecular complex or binding pocket thereof that has a root mean square deviation of conserved residue backbone atoms (N, Cα, C, O) of less than 1.5 Å when superimposed on the relevant backbone atoms described by structure coordinates listed in Figure 1 are considered identical. More preferably, the root mean square deviation is less than 1.0Å.

The term "root mean square deviation" means the square root of the arithmetic mean of the squares of the deviations from the mean. It is a way to express the deviation or variation from a trend or object. For purposes of this invention, the "root mean square deviation" defines the variation in the backbone of a protein from the backbone of NS3 helicase or a binding pocket portion thereof, as defined by the structure coordinates of NS3 helicase described herein.

Therefore, according to another embodiment of this invention is provided a computer for producing:

a) a three-dimensional representation of a molecule or molecular complex, wherein said molecule or molecular complex comprises a binding pocket defined by structure coordinates of NS3 amino acids Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501 and Tyr502 according to Figure 1; or

b) a three-dimensional representation of a homologue of said molecule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å, wherein said computer comprises:

(i) a machine-readable data storage medium comprising a data storage material encoded with machine-readable data, wherein said data comprises the structure coordinates of NS3 amino acids Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501 and Tyr502 according to Figure 1;

(ii) a working memory for storing instructions for processing said machine-readable data;

(iii) a central-processing unit coupled to said working memory and to said machine-readable data storage medium for processing said machine readable data into said three-dimensional representation; and

(iv) a display coupled to said central-processing unit for displaying said three-dimensional representation.

According to a preferred embodiment, the computer produces a three-dimensional representation of:

a) a molecule or molecular complex comprising a binding pocket defined by the structure coordinates of NS3 helicase amino acids Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501, Tyr502, Pro230, Val256, Thr298, Ala497, Lys551, Gln552, Gly554, Glu555, Asn556 and Pro558, according to Figure 1; or

b) a homologue of said molecule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å. In that preferred embodiment, the machine readable data comprises the structure coordinates of NS3 amino acids Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501, Tyr502, Pro230, Val256, Thr298, Ala497, Lys551, Gln552, Gly554, Glu555, Asn556 and Pro558, according to Figure 1.

In the above two embodiments, the computer is producing a three-dimensional graphical structure of a

molecule or a molecular complex which comprises a NS3 helicase U8-like binding pocket.

In an alternate embodiment, the computer produces a three-dimensional representation of:

a) a molecule or molecular complex comprising a binding pocket defined by the structure coordinates of NS3 helicase amino acids His369, Ser370, Lys371, Tyr392, Arg393, Thr411, Asp412, Ala413, Cys431, Val432, Gln434, Ile446, Thr448, Arg461, Glu493, Glu555, Asn556 and Phe557, according to Figure 1; or

b) a homologue of said molecule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å

In this alternate embodiment, the machine readable data comprises the structure coordinates of NS3 amino acids His369, Ser370, Lys371, Tyr392, Arg393, Thr411, Asp412, Ala413, Cys431, Val432, Gln434, Ile446, Thr448, Arg461, Glu493, Glu555, Asn556 and Phe557, according to Figure 1.

In this embodiment, the computer is producing a three-dimensional graphical structure of a molecule or a molecular complex which comprises a NS3 helicase U4-like binding pocket.

In yet another alternate embodiment, the computer produces a three-dimensional representation of:

a) a molecule or molecular complex comprising a binding pocket defined by the structure coordinates of NS3 helicase amino acids Pro205, Thr206, Gly207, Ser208, Gly209, Lys210, Ser211, Thr212, Lys213, Asn229, Ala234, Gly237, Phe238, Tyr241, Asp290, Glu291, His293, Thr322, Ala323, Thr324, Gln460, Gly463, Arg464 and Arg467, according to Figure 1; or

b) a homologue of said molecule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å

In this alternate embodiment, the machine readable data comprises the structure coordinates of NS3 amino acids Pro205, Thr206, Gly207, Ser208, Gly209, Lys210, Ser211, Thr212, Lys213, Asn229, Ala234, Gly237, Phe238, Tyr241, Asp290, Glu291, His293, Thr322, Ala323, Thr324, Gln460, Gly463, Arg464 and Arg467, according to Figure 1.

In this embodiment, the computer is producing a three-dimensional graphical structure of a molecule or a molecular complex which comprises a NS3 helicase NTP-like binding pocket.

Even more preferred is a computer for producing a three-dimensional representation of a molecule or molecular complex defined by structure coordinates of all of the NS3 amino acids set forth in Figure 1, or a three-dimensional representation of a homologue of said molecule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å. In this embodiment, the machined readable data contains the coordinates of all of the NS3

According to an alternate embodiment, the invention provides a computer for determining at least a portion of the structure coordinates corresponding to X-ray diffraction data obtained from a molecule or molecular complex, wherein said computer comprises:

(a) a machine-readable data storage medium comprising a data storage material encoded with machine-readable data, wherein said data comprises at least a portion of the structural coordinates of NS3 helicase according to Figure 1;

(b) a machine-readable data storage medium comprising a data storage material encoded with machine-readable data, wherein said data comprises X-ray diffraction data from said molecule or molecular complex;

(c) a working memory for storing instructions for processing said machine-readable data of (a) and (b);

(d) a central-processing unit coupled to said working memory and to said machine-readable data storage medium of (a) and (b) for performing a Fourier transform of the machine readable data of (a) and for processing said machine readable data of (b) into structure coordinates; and

(e) a display coupled to said central-processing unit for displaying said structure coordinates of said molecule or molecular complex.

For example, the Fourier transform of the structure coordinates set forth in Figure 1 may be used to determine at least a portion of the structure coordinates of other helicases.

Figure 2 demonstrates one version of these embodiments. System 10 includes a computer 11 comprising a central processing unit ("CPU") 20, a working memory 22 which may be, e.g., RAM (random-access memory) or "core" memory, mass storage memory 24 (such as one or more disk drives or CD-ROM drives), one or more cathode-ray tube ("CRT") display terminals 26, one or more keyboards 28, one or more input lines 30, and one or more output lines

40, all of which are interconnected by a conventional bi-directional system bus 50.

Input hardware 36, coupled to computer 11 by input lines 30, may be implemented in a variety of ways. Machine-readable data of this invention may be inputted via the use of a modem or modems 32 connected by a telephone line or dedicated data line 34. Alternatively or additionally, the input hardware 36 may comprise CD-ROM drives or disk drives 24. In conjunction with display terminal 26, keyboard 28 may also be used as an input device.

Output hardware 46, coupled to computer 11 by output lines 40, may similarly be implemented by conventional devices. By way of example, output hardware 46 may include CRT display terminal 26 for displaying a graphical representation of a binding pocket of this invention using a program such as QUANTA as described herein. Output hardware might also include a printer 42, so that hard copy output may be produced, or a disk drive 24, to store system output for later use.

In operation, CPU 20 coordinates the use of the various input and output devices 36, 46, coordinates data accesses from mass storage 24 and accesses to and from working memory 22, and determines the sequence of data

processing steps. A number of programs may be used to process the machine-readable data of this invention. Such programs are discussed in reference to the computational methods of drug discovery as described herein. Specific references to components of the hardware system 10 are included as appropriate throughout the following description of the data storage medium.

Figure 3 shows a cross section of a magnetic data storage medium 100 which can be encoded with a machine-readable data that can be carried out by a system such as system 10 of Figure 2. Medium 100 can be a conventional floppy diskette or hard disk, having a suitable substrate 101, which may be conventional, and a suitable coating 102, which may be conventional, on one or both sides, containing magnetic domains (not visible) whose polarity or orientation can be altered magnetically. Medium 100 may also have an opening (not shown) for receiving the spindle of a disk drive or other data storage device 24.

The magnetic domains of coating 102 of medium 100 are polarized or oriented so as to encode in manner which may be conventional, machine readable data such as that described herein, for execution by a system such as system 10 of Figure 2.

Figure 4 shows a cross section of an optically-readable data storage medium 110 which also can be encoded with such a machine-readable data, or set of instructions, which can be carried out by a system such as system 10 of Figure 2. Medium 110 can be a conventional compact disk read only memory (CD-ROM) or a rewritable medium such as a magneto-optical disk which is optically readable and magneto-optically writable. Medium 100 preferably has a suitable substrate 111, which may be conventional, and a suitable coating 112, which may be conventional, usually of one side of substrate 111.

In the case of CD-ROM, as is well known, coating 112 is reflective and is impressed with a plurality of pits 113 to encode the machine-readable data. The arrangement of pits is read by reflecting laser light off the surface of coating 112. A protective coating 114, which preferably is substantially transparent, is provided on top of coating 112.

In the case of a magneto-optical disk, as is well known, coating 112 has no pits 113, but has a plurality of magnetic domains whose polarity or orientation can be changed magnetically when heated above a certain temperature, as by a laser (not shown). The

orientation of the domains can be read by measuring the polarization of laser light reflected from coating 112. The arrangement of the domains encodes the data as described above.

Thus, in accordance with the present invention, X-ray coordinate data capable of being processed into a three dimensional graphical display of a molecule or molecular complex which comprises a NS3 helicase-like binding pocket is stored in a machine-readable storage medium.

The NS3 helicase X-ray coordinate data, when used in conjunction with a computer programmed with software to translate those coordinates into the 3-dimensional structure of a molecule or molecular complex comprising a NS3 helicase-like binding pocket may be used for a variety of purposes, such as drug discovery.

For example, the structure encoded by the data may be computationally evaluated for its ability to associate with chemical entities. Chemical entities that associate with NS3 helicase may inhibit that enzyme, and are potential drug candidates. Alternatively, the structure encoded by the data may be displayed in a graphical three-dimensional representation on a computer screen. This allows visual inspection of the structure,

as well as visual inspection of the structure's association with chemical entities.

Thus, according to another embodiment, the invention relates to a method for evaluating the potential of a chemical entity to associate with

a) a molecule or molecular complex comprising a binding pocket defined by structure coordinates of NS3 helicase amino acids Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501 and Tyr502 according to Figure 1, or

b) a homologue of said molecule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å.

This method comprises the steps of:

i) employing computational means to perform a fitting operation between the chemical entity and a binding pocket of the molecule or molecular complex; and

ii) analyzing the results of said fitting operation to quantify the association between the chemical entity and the binding pocket.

The term "chemical entity", as used herein, refers to chemical compounds, complexes of at least two chemical compounds, and fragments of such compounds or

complexes.

Preferably, the method evaluates the potential of a chemical entity to associate with

a) a molecule or molecular complex comprising a binding pocket defined by structure coordinates of NS3 helicase amino acids Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501, Tyr502, Pro230, Val256, Thr298, Ala497, Lys551, Gln552, Gly554, Glu555, Asn556 and Pro558 according to Figure 1, or

b) a homologue of said molècule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å.

These embodiments relate to evaluating the potential of a chemical entity to associate with a NS3 helicase U8-like binding pocket.

In an alternate embodiment, the same steps indicated above are used in a method for evaluating the potential of a chemical entity to associate with

a) a molecule or molecular complex comprising a binding pocket defined by structure coordinates of NS3 helicase amino acids His369, Ser370, Lys371, Tyr392, Arg393, Thr411, Asp412, Ala413, Cys431, Val432,

Gln434, Ile446, Thr448, Arg461, Glu493, Glu555, Asn556 and Phe557 according to Figure 1, or

b) a homologue of said molecule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å.

These embodiments relate to evaluating the potential of a chemical entity to associate with a NS3 helicase U4-like binding pocket.

In yet another alternate embodiment, the same steps indicated above are used in a method for evaluating the potential of a chemical entity to associate with

a) a molecule or molecular complex comprising a binding pocket defined by structure coordinates of NS3 helicase amino acids Pro205, Thr206, Gly207, Ser208, Gly209, Lys210, Ser211, Thr212, Lys213, Asn229, Ala234, Gly237, Phe238, Tyr241, Asp290, Glu291, His293, Thr322, Ala323, Thr324, Gln460, Gly463, Arg464 and Arg467 according to Figure 1, or

b) a homologue of said molecule or molecular complex, wherein said homologue comprises a binding pocket that has a root mean square deviation from the

backbone atoms of said amino acids of not more than 1.5Å.

These embodiments relate to evaluating the potential of a chemical entity to associate with a NS3 helicase NTP-like binding pocket.

Even more preferably, the method evaluates the potential of a chemical entity to associate with a molecule or molecular complex defined by structure coordinates of all of the NS3 helicase amino acids, as set forth in Figure 1, or a homologue of said molecule or molecular complex having a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å.

Alternatively, the structural coordinates of the NS3 helicase binding pocket can be utilized in a method for identifying a potential agonist or antagonist of a molecule comprising a NS3 helicase-like binding pocket. This method comprises the steps of:

a. using the atomic coordinates of Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501 and Tyr502 according to Figure 1 ± a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å, to generate a three-dimensional structure of molecule comprising a NS3 helicase-like binding pocket;

b. employing said three-dimensional structure to design or select said potential agonist or antagonist;

c. synthesizing said agonist or antagonist; and

d. contacting said agonist or antagonist with said molecule to determine the ability of said potential agonist or antagonist to interact with said molecule.

More preferred is when the atomic coordinates of Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501, Tyr502, Pro230, Val256, Thr298, Ala497, Lys551, Gln552, Gly554, Glu555, Asn556 and Pro558 according to Figure 1 ± a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å, are used to generate a three-dimensional structure of molecule comprising a NS3 helicase-like binding pocket.

These methods are designed to identify agonists and antagonists that associate with an NS3 helicase U8-like binding pocket.

Alternatively, the atomic coordinates of the NS3 helicase U4 binding pocket -- His369, Ser370, Lys371, Tyr392, Arg393, Thr411, Asp412, Ala413, Cys431, Val432, Gln434, Ile446, Thr448, Arg461, Glu493, Glu555, Asn556 and Phe557 according to Figure 1 -- ± a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å, may be used in step a), above, to generate a three-dimensional structure of molecule comprising a NS3 helicase-like binding pocket.

In another alternative embodiment, the atomic coordinates of the NS3 helicase NTP binding site -- Pro205, Thr206, Gly207, Ser208, Gly209, Lys210, Ser211,

Thr212, Lys213, Asn229, Ala234, Gly237, Phe238, Tyr241, Asp290, Glu291, His293, Thr322, Ala323, Thr324, Gln460, Gly463, Arg464 and Arg467 according to Figure 1 -- ± a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å, may be used in step a), above, to generate a three-dimensional structure of molecule comprising a NS3 helicase-like binding pocket.

Most preferred is when the atomic coordinates of all the amino acids of NS3 helicase according to Figure 1 ± a root mean square deviation from the backbone atoms of said amino acids of not more than 1.5Å, are used to generate a three-dimensional structure of molecule comprising a NS3 helicase-like binding pocket.

For the first time, the present invention permits the use of molecular design techniques to identify, select and design chemical entities, including inhibitory compounds, capable of binding to NS3 helicase-like binding pockets - in particular, the oligonucleotide binding pocket of NS3 helicase.

Applicants' elucidation of the U4 and U8 binding pockets in the oligonucleotide binding site and an expanded elucidation of the NTP binding pocket on NS3 helicase provides the necessary information for designing new chemical entities and compounds that may interact with NS3 helicase-like binding pockets, in whole or in part.

Throughout this section, discussions about the ability of an entity to bind to, associate with or

inhibit a NS3 helicase-like binding pocket refers to features of the entity alone. Assays to determine if a compound binds to NS3 helicase are well known in the art and are exemplified below.

The design of compounds that bind to or inhibit NS3 helicase-like binding pockets according to this invention generally involves consideration of two factors. First, the entity must be capable of physically and structurally associating with parts or all of the NS3 helicase-like binding pockets. Non-covalent molecular interactions important in this association include hydrogen bonding, van der Waals interactions, hydrophobic interactions and electrostatic interactions.

Second, the entity must be able to assume a conformation that allows it to associate with the NS3 helicase-like binding pocket directly. Although certain portions of the entity will not directly participate in these associations, those portions of the entity may still influence the overall conformation of the molecule. This, in turn, may have a significant impact on potency. Such conformational requirements include the overall three-dimensional structure and orientation of the chemical entity in relation to all or a portion of the binding pocket, or the spacing between functional groups

of an entity comprising several chemical entities that directly interact with the NS3 helicase-like binding pocket or homologues thereof.

The potential inhibitory or binding effect of a chemical entity on a NS3 helicase-like binding pocket may be analyzed prior to its actual synthesis and testing by the use of computer modeling techniques. If the theoretical structure of the given entity suggests insufficient interaction and association between it and the NS3 helicase-like binding pocket, testing of the entity is obviated. However, if computer modeling indicates a strong interaction, the molecule may then be synthesized and tested for its ability to bind to a NS3 helicase-like binding pocket. This may be achieved by testing the ability of the molecule to inhibit NS3 helicase using the assays described in Example 5. In this manner, synthesis of inoperative compounds may be avoided.

A potential inhibitor of a NS3 helicase-like binding pocket may be computationally evaluated by means of a series of steps in which chemical entities or fragments are screened and selected for their ability to associate with the NS3 helicase-like binding pockets.

One skilled in the art may use one of several methods to screen chemical entities or fragments for their ability to associate with a NS3 helicase-like binding pocket. This process may begin by visual inspection of, for example, a NS3 helicase-like binding pocket on the computer screen based on the NS3 helicase structure coordinates in Figure 1 or other coordinates which define a similar shape generated from the machine-readable storage medium. Selected fragments or chemical entities may then be positioned in a variety of orientations, or docked, within that binding pocket as defined supra. Docking may be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics with standard molecular mechanics force fields, such as CHARMM and AMBER.

Specialized computer programs may also assist in the process of selecting fragments or chemical entities. These include:

1. GRID (P. J. Goodford, "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", J. Med. Chem., 28, pp. 849-857 (1985)). GRID is available from Oxford University, Oxford, UK.

2. MCSS (A. Miranker et al., "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method." Proteins: Structure, Function and Genetics, 11, pp. 29-34 (1991)). MCSS is available from Molecular Simulations, San Diego, CA.

3. AUTODOCK (D. S. Goodsell et al., "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8, pp. 195-202 (1990)). AUTODOCK is available from Scripps Research Institute, La Jolla, CA.

4. DOCK (I. D. Kuntz et al., "A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161, pp. 269-288 (1982)). DOCK is available from University of California, San Francisco, CA.

Once suitable chemical entities or fragments have been selected, they can be assembled into a single compound or complex. Assembly may be preceded by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of NS3 helicase. This would be followed by manual model

building using software such as Quanta or Sybyl [Tripos Associates, St. Louis, MO].

Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include:

1. CAVEAT (P. A. Bartlett et al, "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules", in Molecular Recognition in Chemical and Biological Problems", Special Pub., Royal Chem. Soc., 78, pp. 182-196 (1989); G. Lauri and P. A. Bartlett, "CAVEAT: a Program to Facilitate the Design of Organic Molecules", J. Comput. Aided Mol. Des. , 8, pp. 51-66 (1994)). CAVEAT is available from the University of California, Berkeley, CA.

2. 3D Database systems such as ISIS (MDL Information Systems, San Leandro, CA). This area is reviewed in Y. C. Martin, "3D Database Searching in Drug Design", J. Med. Chem., 35, pp. 2145-2154 (1992).

3. HOOK (M. B. Eisen et al, "HOOK: A Program for Finding Novel Molecular Architectures that Satisfy the Chemical and Steric Requirements of a Macromolecule Binding Site", Proteins: Struct., Funct., Genet., 19,

pp. 199-221 (1994). HOOK is available from Molecular Simulations, San Diego, CA.

Instead of proceeding to build an inhibitor of a NS3 helicase-like binding pocket in a step-wise fashion one fragment or chemical entity at a time as described above, inhibitory or other NS3 helicase binding compounds may be designed as a whole or "de novo" using either an empty binding site or optionally including some portion(s) of a known inhibitor(s). There are many de novo ligand design methods including:

1. LUDI (H.-J. Bohm, "The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors", J. Comp. Aid. Molec. Design, 6, pp. 61-78 (1992)). LUDI is available from Molecular Simulations Incorporated, San Diego, CA.

2. LEGEND (Y. Nishibata et al., Tetrahedron, 47, p. 8985 (1991)). LEGEND is available from Molecular Simulations Incorporated, San Diego, CA.

3. LeapFrog (available from Tripos Associates, St. Louis, MO).

4. SPROUT (V. Gillet et al, "SPROUT: A Program for Structure Generation)", J. Comput. Aided Mol. Design, 7,

pp. 127-153 (1993)). SPROUT is available from the University of Leeds, UK.

Other molecular modeling techniques may also be employed in accordance with this invention [see, e.g., N. C. Cohen et al., "Molecular Modeling Software and Methods for Medicinal Chemistry, J. Med. Chem., 33, pp. 883-894 (1990); see also, M. A. Navia and M. A. Murcko, "The Use of Structural Information in Drug Design", Current Opinions in Structural Biology, 2, pp. 202-210 (1992); L. M. Balbes et al., "A Perspective of Modern Methods in Computer-Aided Drug Design", in Reviews in Computational Chemistry, Vol. 5, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, pp. 337-380 (1994); see also, W. C. Guida, "Software For Structure-Based Drug Design", Curr. Opin. Struct. Biology,, 4, pp. 777-781 (1994)].

Once a compound has been designed or selected by the above methods, the efficiency with which that entity may bind to an NS3 helicase binding pocket may be tested and optimized by computational evaluation. For example, an effective NS3 helicase binding pocket inhibitor must preferably demonstrate a relatively small difference in energy between its bound and free states (i.e., a small deformation energy of binding). Thus, the

most efficient NS3 helicase binding pocket inhibitors should preferably be designed with a deformation energy of binding of not greater than about 10 kcal/mole, more preferably, not greater than 7 kcal/mole. NS3 helicase binding pocket inhibitors may interact with the binding pocket in more than one conformation that is similar in overall binding energy. In those cases, the deformation energy of binding is taken to be the difference between the energy of the free entity and the average energy of the conformations observed when the inhibitor binds to the protein.

An entity designed or selected as binding to an NS3 helicase binding pocket may be further computationally optimized so that in its bound state it would preferably lack repulsive electrostatic interaction with the target enzyme and with the surrounding water molecules. Such non-complementary electrostatic interactions include repulsive charge-charge, dipole-dipole and charge-dipole interactions.

Specific computer software is available in the art to evaluate compound deformation energy and electrostatic interactions. Examples of programs designed for such uses include: Gaussian 94, revision C (M. J. Frisch, Gaussian, Inc., Pittsburgh, PA ©1995);

AMBER, version 4.1 (P. A. Kollman, University of California at San Francisco, ©1995); QUANTA/CHARMM (Molecular Simulations, Inc., San Diego, CA ©1995); Insight II/Discover (Molecular Simulations, Inc., San Diego, CA ©1995); DelPhi (Molecular Simulations, Inc., San Diego, CA ©1995); and AMSOL (Quantum Chemistry Program Exchange, Indiana University). These programs may be implemented, for instance, using a Silicon Graphics workstation such as an Indigo^2 with "IMPACT" graphics. Other hardware systems and software packages will be known to those skilled in the art.

Another approach enabled by this invention, is the computational screening of small molecule databases for chemical entities or compounds that can bind in whole, or in part, to a NS3 helicase binding pocket. In this screening, the quality of fit of such entities to the binding site may be judged either by shape complementarity or by estimated interaction energy [E. C. Meng et al., J. Comp. Chem., 13, pp. 505-524 (1992)].

According to another embodiment, the invention provides compounds which associate with a NS3 helicase-like binding pocket produced or identified by the method set forth above.

The structure coordinates set forth in Figure 1 can also be used to aid in obtaining structural information about another crystallized molecule or molecular complex. This may be achieved by any of a number of well-known techniques, including molecular replacement.

Therefore, in another embodiment this invention provides a method of utilizing molecular replacement to obtain structural information about a molecule or molecular complex whose structure is unknown comprising the steps of:

a) crystallizing said molecule or molecular complex of unknown structure;

b) generating X-ray diffraction data from said crystallized molecule or molecular complex; and

c) applying at least a portion of the structure coordinates set forth in Figure 1 to the X-ray diffraction data to generate a three-dimensional electron density map of the molecule or molecular complex whose structure is unknown.

By using molecular replacement, all or part of the structure coordinates of the NS3 helicase/ oligonucleotide complex as provided by this invention (and set forth in Figure 1) can be used to determine the

structure of a crystallized molecule or molecular complex whose structure is unknown more quickly and efficiently than attempting to determine such information ab initio.

Molecular replacement provides an accurate estimation of the phases for an unknown structure. Phases are a factor in equations used to solve crystal structures that can not be determined directly. Obtaining accurate values for the phases, by methods other than molecular replacement, is a time-consuming process that involves iterative cycles of approximations and refinements and greatly hinders the solution of crystal structures. However, when the crystal structure of a protein containing at least a homologous portion has been solved, the phases from the known structure provide a satisfactory estimate of the phases for the unknown structure.

Thus, this method involves generating a preliminary model of a molecule or molecular complex whose structure coordinates are unknown, by orienting and positioning the relevant portion of the NS3 helicase/oligonucleotide complex according to Figure 1 within the unit cell of the crystal of the unknown molecule or molecular complex so as best to account for the observed X-ray diffraction data of the crystal of the

molecule or molecular complex whose structure is unknown. Phases can then be calculated from this model and combined with the observed X-ray diffraction data amplitudes to generate an electron density map of the structure whose coordinates are unknown. This, in turn, can be subjected to any well-known model building and structure refinement techniques to provide a final, accurate structure of the unknown crystallized molecule or molecular complex [E. Lattman, "Use of the Rotation and Translation Functions", in Meth. Enzymol., 115, pp. 55-77 (1985); M. G. Rossmann, ed., "The Molecular Replacement Method", Int. Sci. Rev. Ser., No. 13, Gordon & Breach, New York (1972)].

The structure of any portion of any crystallized molecule or molecular complex that is sufficiently homologous to any portion of the NS3 helicase/oligonucleotide complex can be resolved by this method.

In a preferred embodiment, the method of molecular replacement is utilized to obtain structural information about another helicase. The structure coordinates of NS3 helicase as provided by this invention are particularly useful in solving the structure of other

isoforms of NS3 helicase or other NS3 helicase-containing complexes.

Furthermore, the structure coordinates of NS3 helicase as provided by this invention are useful in solving the structure of NS3 helicase proteins that have amino acid substitutions, additions and/or deletions (referred to collectively as "NS3 helicase mutants", as compared to naturally occurring NS3 helicase isoforms. These NS3 helicase mutants may optionally be crystallized in co-complex with a chemical entity, such as a non-hydrolyzable NTP analogue or an oligonucleotide. The crystal structures of a series of such complexes may then be solved by molecular replacement and compared with that of wild-type NS3 helicase. Potential sites for modification within the various binding sites of the enzyme may thus be identified. This information provides an additional tool for determining the most efficient binding interactions, for example, increased hydrophobic interactions, between NS3 helicase and a chemical entity or compound.

The structure coordinates are also particularly useful to solve the structure of crystals of NS3 helicase or NS3 helicase homologues co-complexed with a variety of chemical entities. This approach enables the

determination of the optimal sites for interaction between chemical entities, including between candidate NS3 helicase inhibitors and NS3 helicase. For example, high resolution X-ray diffraction data collected from crystals exposed to different types of solvent allows the determination of where each type of solvent molecule resides. Small molecules that bind tightly to those sites can then be designed and synthesized and tested for their NS3 helicase inhibition activity.

All of the complexes referred to above may be studied using well-known X-ray diffraction techniques and may be refined versus 1.5-3 Å resolution X-ray data to an R value of about 0.20 or less using computer software, such as X-PLOR [Yale University, ©1992, distributed by Molecular Simulations, Inc.; see, e.g., Blundell & Johnson, supra; Meth. Enzymol., vol. 114 & 115, H. W. Wyckoff et al., eds., Academic Press (1985)]. This information may thus be used to optimize known NS3 helicase inhibitors, and more importantly, to design new NS3 helicase inhibitors.

In order that this invention be more fully understood, the following examples are set forth. These examples are for the purpose of illustration only and are

not to be construed as limiting the scope of the invention in any way.

EXAMPLE 1

Cloning and Expression of NS3 Helicase

The HCV NS3 RNA helicase domain (encoded by nucleotides 502-1896 of SEQ ID NO:6) was subcloned from a cDNA of the HCV H strain [A. Grakoui et al., J. Virol., 67, pp. 1385-95 (1993); C. Lin et al., J. Virol., 68, pp. 8147-57 (1994), the disclosures of which are herein incorporated be reference] into a pET expression vector (Novagen, Madison, WI). The resulting plasmid, pET-BS(+)/HCV/NS3-C465-His (SEQ ID NO:4), also contained a methionine start codon, a linker encoded Gly-Ser-Gly-Ser sequence attached the C-terminal threonine of the NS3 helicase domain and a six-histidine tag fused to the C-terminus of the Gly-Ser-Gly-Ser sequence to facilitate protein purification. This plasmid was used as a template for single-stranded DNA-based site-directed mutagenesis essentially as described by (T.A. Kunkel, Proc. Natl. Acad. Sci. USA, 82, pp.488-492 (1985) and C. Lin et al., Virology, 192, pp.596-604 (1993), the disclosures of which are herein incorporated by reference) with the following modifications.

The single stranded phagemid DNA packaged in the presence of helper M13 phage corresponds to the HCV plus strand. A single colony of *E. coli* strain CJ 326, transformed with pET-BS(+)/HCV/NS3-C465-His, was grown in YT media containing 0.25 μg/ml of uridine and 50 μg/ml of carbenicillin. After three serial passages, M13 helper phage (Bio-Rad) was used to rescue uridylated phagemid single stranded DNA, which was then used as template for oligonucleotide-directed mutagenesis [T.A. Kunkel (1985), supra]. ABI automatic sequencing was used to identify mutations and ensure that there is no other unintended mutation within the HCV NS3 helicase domain sequences. Construct containing mutations were named according to the position of the substituted residue in the full-length HCV NS3 protein.

In this manner, we made NS3 helicase corresponding to the consensus sequence of the HCV genotype 1 (Pro at amino acid 332; Ser at amino acid 403; Ala at amino acid 410; and Thr at amino acid 505; hereinafter referred to as "wild type"); as well as NS3 helicase containing the following single amino acid mutations as compared to the consensus HCV genotype 1 NS3 helicase sequence: Ser231-to-Ala; Thr269-to-Ala; Ser370-to-Ala; Thr411-to-Ala; Trp501-to-Phe; Trp501-to-Leu; and

Trp501-to-Ala.

E. coli BL21(DE3) cells, freshly transformed with the pET-BS(+)/HCV/NS3-C465-His plasmid or similar plasmids encoding the single amino acid mutant NS3 helicases described above, were grown at 30°C in LB media supplemented with 50 µg/ml of carbenicillin. When the density reached an OD_{600} of 1.0, the cells were induced for 3 hr at 30°C by the addition of IPTG to a final concentration of 0.8 mM. After induction, the cells were harvested and stored frozen at -70°C until purification.

All the protein purification procedures were performed at 4°C. Typically, 10 g of cell paste was resuspended in 50 ml of buffer A [50 mM HEPES (pH 8), 300 mM NaCl, 10 % glycerol, and 2.5 mM β-mercaptoethanol] containing 0.2 mM phenylmethylsulfonyl fluoride (PMSF), and lysed using a microfluidizer. The lysate was clarified by centrifugation at 100,000 x g for 35 min. We then added 5 mM imidazole (pH 8) to the supernatant and the resulting solution was incubated for 2 hours with 2 ml of Ni-NTA-agarose (Qiagen, Chatsworth, CA).

The resin was packed into a column and washed with 10 bed volumes of buffer A containing 5 mM and 15 mM imidazole, and eluted with buffer A containing 100 mM imidazole. The eluant was desalted to buffer B [50 mM

HEPES (pH 8), 10 % glycerol, and 2.5 mM β-mercaptoethanol] containing 50 mM NaCl on a PD-10 column (Pharmacia). The desalted solution was loaded onto a Heparin-Sepharose column (Pharmacia). The flow-through was then applied onto a Q-Sepharose column (Pharmacia) and washed with 10 bed volumes of buffer B containing 50 mM NaCl. The column was then eluted with a NaCl gradient from 50 mM to 2M in buffer B.

The peak fraction containing the HCV NS3 helicase domain protein was shown by gel-filtration chromatography to be monomeric. The purified protein was judged to be greater than 90% pure by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) and Coomassie R-250 staining.

For crystallization studies, the protein was concentrated to 10 mg/ml by ultrafiltration and slowly diluted with 5 volumes of 15 mM MES (pH 6.5), 2.5 mM β-mercaptoethanol and again concentrated to 10 mg/ml. The dilution step was then repeated with 2 volumes of the MES buffer and concentrated to 13 mg/ml. We then added oligonucleotide (dU_8; Oligo Therapeutics, Inc., Wilsonville, OR) to yield a 1:1 molar ration of protein to nucleic acid.

To produce recombinant full-length NS3 protein which was subsequently used for mutagenesis studies in the NTP binding pocket, we followed similar procedures as above. The full length NS3 coding sequence was also subcloned from the HCV H strain. It was placed in a pET expression vector to create the plasmid pET-BS(+)/HCV/FLNS3-His (SEQ ID NO:3). As with the helicase constructs previously described, the full-length NS3 coding sequence was preceded by a methionine start codon and had codons encoding Gly-Ser-Gly-Ser-His_6 in frame at the C-terminus of the NS3 coding region. That plasmid was used as a template for single-stranded DNA-based site-directed mutagenesis as described above.

In this manner, we made NS3 containing the following single amino acid mutations as compared to the consensus HCV genotype 1 NS3 helicase sequence: Gln460->Ala, Arg461->Ala, Arg462->Ala, Arg464->Ala and Arg467->Ala.

Both the wild-type full-length NS3 and the single amino acid mutants were purified as described above.

EXAMPLE 2

Crystallization and Data Collection

Crystals of the NS3 helicase:dU8 complex were grown by hanging-drop vapor diffusion over wells containing 0.1 M Tris pH 8.0, 0.2 M Li_2SO_4, 18% Polyethylene glycol 6000, and 8 mM β-mercaptoethanol. Drops were macroseeded within 12 hours after being set up. Crystals grew over the course of 2-3 weeks to dimensions of 0.4 x 0.4 x 0.1 mm^3. The crystals belong to space group $P2_12_12$ with unit cell dimensions a=73.1Å b=117.5Å, c=63.4Å, and contain one helicase:dU8 complex per asymmetric unit.

Heavy atom soaks were carried out by transferring crystals to a solution containing 0.1 M Tris pH 8.0, 0.2 M Li_2SO_4, 17% Polyethylene glycol 6000, 8 mM β-mercaptoethanol, in addition to the heavy atom in question. Heavy atom soaks with K_2WO_4 were performed in the absence of Li_2SO_4.

Crystals were transferred to a solution containing 0.08 M Tris pH 8.0, 0.2 M Li_2SO_4, 16% Polyethylene glycol 6000, 8 mM β-mercaptoethanol, and 15% glycerol and immediately frozen in a dry nitrogen gas stream at 100 K (Molecular Structure Corp., Houston, TX) for data collection.

Data was acquired by oscillation photography on a Rigaku R-AXIS IIC phosphor imaging area detector

mounted on a Rigaku RU200 rotating anode generator (MSC), operating at 50kV and 100mA. Measured intensities were integrated, scaled, and merged using the HKL software package [Z. Otwinowski et al., Meth. Enzymol., 276, pp. 307-326 (1997)].

EXAMPLE 3

Phasing, model building and refinement

Heavy atom positions were located from difference Patterson and anomalous difference Patterson maps and confirmed with difference Fourier syntheses. Heavy atom parameters were refined and phases computed to 2.3Å using the program PHASES [W. Furey et al., Meth. Enzymol., 277, pp. 590-620, (1997) need full cite]. MIR phases were improved by cycles of solvent flattening [B. C. Wang,. Methods Enzymol., 115, pp. 90-112 (1985)] combined with histogram matching [K. Y. J. Zhang et al., Acta Crystallogr., A46, pp.377-381 (1990)] using the CCP4 crystallographic package [CCP4; C. C. Project, Acta Crystallogr., D50, pp. 760-763 (1994)].

Model building was carried out using QUANTA96 (Molecular Simulations), and all refinement done in XPLOR [A. T. Brunger, "X-PLOR: A System for X-Ray Crystallography and NMR," Yale University Press, New

Haven, Connecticut. (1993)], using the free R-value [A. T. Brunger, Nature, 355, pp. 472-475 (1992)] to monitor the course of refinement. The current model, refined using data from 6.0-2.2Å, consists of NS3 helicase residues 190-414 and 418-626, residues 3-8 of dU_8, 1 bound sulfate ion, and 159 well-ordered water molecules.

EXAMPLE 4

Structural Features of the NS3 Helicase-dU_8 Complex

The structure of the resolved portion of HCV NS3 helicase (NS3 residues 189-626 of SEQ ID NO:1, corresponding to HCV polyprotein residues 1215-1652) complexed with a deoxyuridine octamer (dU_8) was determined by multiple isomorphous replacement combined with anomalous scattering. The protein consists of three domains separated by a series of clefts (Figure 5).

Domains 1 and 3 share a more extensive interface than either share with domain 2. This interface is largely accounted for by packing of helices α5 and α6 from domain 3 on helix α4 from domain 1. As a result, the clefts between domains 1 and 2 and domains 2 and 3 are the largest. A published crystal structure of HCV NS3 helicase domain demonstrated that domain 2 could undergo rigid body movements relative to domains 1 and 3

based on a comparison of two crystallographically independent molecules [N. Yao et al., Nat. Struct. Biol., 4, pp. 463-467 (1997)]. Preliminary structural studies on HCV helicase in a different crystal form also show a rotation of domain 2 relative to the first and third, confirming that this domain is flexibly linked to the other two.

Domains 1 and 2 of the HCV helicase, which contain all of the conserved helicase sequence motifs, have similar topologies (Figure 6) and are also similar to domains 1A and 2A of the four domain PcrA and Rep DNA helicases [H. S. Subramanya et al., Nature, 384, pp. 379-383 (1996); S. Korolev et al., Cell, 90, pp. 635-647 (1997)]. The structurally homologous domains of PcrA, Rep, and HCV helicases each contain a parallel six-stranded ß-sheet flanked by α-helices. In addition, domain 1 of HCV helicase contains a seventh ß-strand running antiparallel to the rest of the sheet.

Superposition of domains 1 and 2 yields an rms deviation of 2.0 Å for 76 C-alpha atoms that form the core of each domain. Domain 3 is predominantly α-helical and is associated with domain 2 by a pair of anti-parallel ß-strands (Figure 5). An interesting component of domain 3 is a 40 amino acid region preceding the C-

terminal α-helix that lacks secondary structure elements. This may represent a flexible region of the protein that allows the C-terminus of NS3 to reach the active site of its own serine protease domain to facilitate cleavage at the NS3/NS4A junction during HCV polyprotein processing. This cleavage is believed to occur in cis [R. Bartenschlager et al., J. Virol., 68, pp. 5045-5055 (1994)].

The N-terminal region of domain 1 contains a phosphate binding loop that is highly conserved among all helicases and commonly referred to as the Walker A box or motif I [J. E. Walker et al., EMBO J., 1, pp. 945-951 (1982)]. In the structure presented here, this loop contains a bound sulfate ion (Figure 7A). This phosphate binding loop is structurally similar to those found in a number of other ATPases [M. Saraste et al., Trends Biochem. Sci., 15, 430-434 (1990)]. The sulfate ion is stabilized by hydrogen bonds from the amide nitrogens of Gly-207 and Gly-209, and the side chains of Ser-208, Lys-210, and Ser-211. Lys-210 makes an additional water-mediated contact to the conserved Asp-290 of the DECH motif (motif II or Walker motif B).

The side chains of Asp-290 and Glu-291, the most conserved residues in the DECH motif, point toward

an open area beneath the phosphate binding loop that presumably is occupied by Mg^{2+} and the γ-phosphate of the bound Mg^{2+}-NTP substrate. Cys-292 is buried in the protein interior while the His-293 side chain points into the cleft between domains 1 and 2. The position of the sulfate in this structure appears to be very similar to that of the ß-phosphate of ADP in the PcrA helicase:ADP complex [H. S. Subramanya et al., Nature, (1996), supra]. It is therefore likely that this sulfate ion occupies the position of the ß-phosphate when NTP or NDP is bound to the HCV helicase.

Highly conserved residues Gln-460, Arg-464, and Arg-467 from domain 2 are solvent exposed in the interdomain cleft, while Arg-461 and Arg-462 are buried in domain 2 and stabilized by internal salt bridges and hydrogen bonds. The position of Arg-461 contrasts that described in the structure of the apo HCV NS3 helicase, which reported this side chain as being solvent exposed and interacting with the phosphate backbone of single stranded nucleic acid modeled into this cleft [N. Yao et al., Nat. Struct. Biol., (1997), supra].

Protein-single stranded DNA interactions

Studies of HCV NS3 helicase single stranded nucleic acid binding have demonstrated that poly(dU) binds to the helicase with higher affinity than poly(rU) of the same length [F. Preugschat et al, J. Biol. Chem., 271 (1996), supra]. Extrapolation of this data suggested that a deoxyoligonucleotide 8mer might be long enough to bind to the helicase with high affinity and not interfere with protein-protein contacts during crystallization. Therefore an oligo dU_8 was used for complex crystallization. In the structure presented here the first two residues of the oligonucleotide are disordered and have not been included in the model. The sugar-phosphate backbone of the third nucleotide is well represented in electron density maps (Figure 7B) while density for the base is extremely weak. Electron density for residues 4-8 is very well defined for the sugar-phosphate backbone and slightly weaker for the bases. Preliminary studies with dU_{10} and dU_{12} oligonucleotides show essentially the same electron density for the DNA.

The bound single stranded DNA ("ssDNA") lies in a channel approximately 16 angstroms in diameter that separates domain 3 from domains 1 and 2 (Figure 10). The 5' end of the oligonucleotide resides at the interface of domains 2 and 3 and its 3' end at the interface of

domains 1 and 3. This orientation of the DNA is roughly perpendicular to that of the ssRNA in a model derived from the apo HCV helicase structure, in which the ssRNA was placed in the cleft between domains 1 and 2 [N. Yao et al., Nat. Struct. Biol., (1997), supra]. It is, however, consistent with the oligonucleotide binding site in the Rep helicase:DNA structure [S. Korolev et al., Cell, (1997), supra].

Interactions between the ssDNA and enzyme are mostly confined to the DNA backbone, as would be expected for a nonspecific protein-nucleic acid complex, and are concentrated at the two ends of the oligonucleotide. Protein contacts emanate mostly from loops between secondary structural elements in domains 1 and 2 (Figures 6A,B). Interestingly, these contacts arise from symmetrically equivalent residues in these two domains, so that protein contacts to the dU4 and dU5 backbone phosphates are nearly identical to those to the dU7 and dU8 phosphates.

At the 3' end of the DNA the dU8 phosphate is stabilized by a hydrogen bond with Thr-269 Oγ, which in turn accepts a hydrogen bond from the main chain NH of Lys-272, and by a hydrogen bond to the main chain NH of Gly-255. Equivalent contacts to the dU5 phosphate are

made by the Arg-393 main chain NH and Thr-411 Oγ, which accepts a hydrogen bond from the Ala-413 NH. The dU7 phosphate accepts a hydrogen bond from the Val-232 NH and interacts with the Ala-233 NH and Ser-231 Oγ via a bridging water molecule. The direct and water mediated main chain interactions are duplicated by Lys-371 and Lys-372 from domain 2 to the dU4 phosphate. Ser-370, the equivalent residue in domain 2 to Ser-231, makes a water mediated contact to the dU3 phosphate rather than dU4. Superposition of domains 1 and 2 of HCV helicase reveals that the residues involved in phosphate contacts are structurally equivalent (Figure 6). This was an unanticipated finding based on the poor sequence homology between these two domains. Additionally, the four residues interacting with the phosphate backbone, Ser-231, Thr-269, Ser-370, and Thr-411, are absolutely conserved in all HCV NS3 sequences known to date. These findings suggest that these two domains may have arisen from a gene duplication event.

Residues dU4-dU8 are capped by interactions at each end with hydrophobic side chains. Trp-501 stacks with the base of dU8 while Val-432 interacts with the dU4 base (Figure 10). These two side chains act as a pair of bookends, defining a central binding cavity occupied by

five nucleotides. Both Val-432 and Trp-501 are highly conserved among HCV NS3 sequences but neither have been implicated in nucleic acid binding nor duplex unwinding. The Val-432:dU4 base interaction induces significant rotation about the phosphate backbone between dU3 and dU4 such that the bases are completely unstacked (Figure 10). Stacking of Trp-501 with the dU8 base should similarly necessitate a large rotation about the phosphate of the following nucleotide. The resulting conformation of the DNA could be stabilized by phosphate interactions with Arg-253 and Lys-272 from domain 1 and Lys-372 and Lys-373 from domain 2, which lie outside the central binding cavity.

Domain 2 contains a pair of extended anti-parallel strands encompassing residues 430-452, which are involved in binding the 5' end of the oligonucleotide (Figure 5). Two other single-strand polynucleotide binding proteins, SSB and tRNA synthetase, contain anti-parallel strands extending from their protein core that are thought to make up the nucleic acid binding site [S. Raghunathan et al., Proc. Natl. Acad. Sci. USA, 94, pp. 6652-6657 (1997); M. Ruff et al., Science, 252, pp. 1682-1689 (1991)]. This region is termed the L_{45} loop in this class of nucleic acid binding proteins. In the HCV

helicase structure the oligonucleotide binds in a channel spanning two protein domains in a manner roughly similar to that seen for replication protein A (RPA) [A. Bochkarev et al., Nature, 385, pp. 176-181 (1997)]. In both structures, the oligonucleotide is most tightly bound at the 3' and 5' ends with few contacts with the central nucleotides. RPA also contains an L_{45} loop, which binds to the 5' end of the oligonucleotide.

The five residues occupying the central binding cavity of HCV helicase adopt a conformation reminiscent of the central base pairs of DNA in the TBP-TATA box complex structures [Y. Kim et al., Nature, 365, pp. 512-520 (1993); J. L. Kim et al., Nature, 365, pp. 520-527 (1993)]. In both instances the DNA is underwound considerably and the backbone smoothly bent, compressing the edges of the bases. Comparison of the DNA structure here with the central pyrimidine stretch of the TATA box DNA reveals that this DNA is more underwound than that seen in the TBP-TATA box complex.

Our structure of the helicase:dU_8 complex does not offer a ready explanation as to why the enzyme binds to poly(dU) with higher affinity than to other homopolymer DNAs [F. Preugschat et al, J. Biol. Chem., 271 (1996), supra]. Sequence specific interactions with

the DNA bases are not observed within the central binding cavity of the helicase. Differences in DNA binding affinity between different sequences in this case may be a result of differences in energetics of DNA distortion and base stacking rather than base-specific hydrogen bonding patterns.

Location of conserved sequence motifs

There is very high sequence conservation among various HCV strains in the NS3 RNA helicase domain with >80% sequence identity over the entire 456 amino acid polypeptide. The most highly conserved segments of these domains correspond to the canonical helicase sequence motifs (Figures 8, 9A) [A. Gorbalenya et al., Curr. Opin. Struct. Biol., 3, pp. 419-429 (1993)]. In the three-dimensional structure residues from these motifs form the interface between the first two domains (Figure 9B). Visual inspection of the structures of the PcrA DNA helicase from *Bacillus stearothermophilus* [H. S. Subramanya et al., Nature, (1996), supra] and *E. coli* Rep DNA helicase [S. Korolev et al., Cell (1997), supra; coordinates not available] suggests overall structural similarity between domains 1A and 2A of these DNA helicases and domains 1 and 2 of the HCV helicase. The

locations of the conserved DNA helicase sequence motifs overlap with those of the HCV helicase allowing an unambiguous alignment of these motifs. Mutagenesis of individual residues within these motifs in HCV helicase or in other RNA helicases have demonstrated that they are essential for enzyme activity. The individual phenotypes of these mutants can now be more fully explained using the enzyme structure.

Domain 1 of the HCV helicase has a fold similar to that found in a number of adenosine triphosphate transphosphorylases, such as adenylate and thymidine kinases. In particular, the phosphate binding loop formed by motif I (GSGKT) is virtually identical to the corresponding loop in the kinases. In these kinases this loop is involved in binding the ß phosphate of ATP. HCV helicase has a sulfate bound in this exact location (Figure 7A). Mutation of residues corresponding to HCV helicase Lys-210 in other helicases invariably leads to inactivation [J. W. George et al., J. Mol. Biol., 235, 424-435 (1994); T. W. Seeley et al., J. Biol. Chem., 265, pp. 7158-7165 (1990)].

Motif II (DExH) is proximal to the GSGKT phosphate binding loop and is expected to be involved in binding the Mg^{2+}-ATP substrate. In adenylate and

thymidine kinases, a conserved aspartate binds Mg^{2+}, which helps orient the ATP for nucleophilic attack [M. E. Black et al., J. Biol. Chem., 267, pp. 6801-6806 (1992); H. G. Yan et al., Biochemistry, 30, pp. 5539-5546 (1991)]. Mutation of the equivalent aspartate residue in these kinases or in other helicases inactivates ATP hydrolysis [M. E. Black et al., J. Biol. Chem., (1992), supra; C. H. Gross et al., J. Virol., 69, pp. 4727-4736 (1995); R. M. Brosh Jr., et al., J. Bacteriol., 177, pp. 5612-562 (1995); A. Pause et al., EMBO J., 11, pp. 2643-2654 (1992)]. His-293 is located at the bottom of the interdomain cleft and approximately 4 Å away from Val-456 and Gln-460. This histidine appears to be essential for coupling the ATPase activity to polynucleotide binding; mutations of this histidine in HCV NS3 and vaccinia NPH-II helicases result in a functional ATPase with no helicase activity [C. H. Gross et al., J. Virol., (1995), supra; G. M. Heilek et al. J. Virol., 71, 6264-6266 (1997)]. Unfortunately, the structure presented here does not provide an obvious explanation as to how this residue couples the NTPase and unwinding activities.

Studies in several helicases have looked at the effects of mutations in motif VI (QRxGRxGR), yet a role for this motif has not been clearly defined. In the HCV

helicase residues in this motif are located in the 1:2 interdomain cleft. Gln-460 lies at the bottom of the cleft opposite from His-293. Mutation of the corresponding glutamine in vaccinia virus helicase and in eIF-4A leads to significant decreases in ATPase activity [A. Pause et al., EMBO J., (1992), supra; C. H. Gross et al., J. Virol., 70, pp. 1706-1713 (1996)]. There are three conserved arginines in motif VI that were proposed by N. Yao et al., Nat. Struct. Biol. (1997), supra, to be involved in binding single-stranded RNA in the cleft between domains 1 and 2. Our structure of the helicase:dU_8 complex demonstrates that this interpretation is unlikely to be correct.

Here Arg-461 points away from the cleft and is hydrogen-bonded to Asp-412 and Asp-427. Mutation of this residue in a vaccinia virus helicase leads to decreases in RNA binding [C. H. Gross et al., J. Virol., (1996), supra], possibly as a consequence of alterations in the conformation of Asp-412 which lines the polynucleotide binding channel. Arginines 464 and 467 extend into the interdomain cleft, directly across from the presumed locations of the γ and α phosphates of ATP. These residues appear to be poised to contact the ATP phosphates upon closure of this interdomain cleft. This

would be similar to the function of conserved basic residues in the second domain of adenylate kinase.

Consistent with the possibility that Arg-464 and Arg-467 are directly involved in ATP binding, mutations of the corresponding residues to Ala or Gln in vaccinia NPH-II or eIF-4A reduce the ATPase activity to <20% of wild type levels [C. H. Gross et al., J. Virol., (1996), supra; A. Pause et al., Mol. Cell. Biol., 13, pp. 6789-6798 (1993)]. Arg-467 appears to be conserved among all three helicase superfamilies (Figure 9A) [A. Gorbalenya et al., Curr. Opin. Struct. Biol. (1993), supra]

Motif III connects domains 1 and 2, and appears to be a flexible linker [N. Yao et al., Nat. Struct. Biol. (1997), supra]. Motif Ia forms part of the ß sheet core of domain 1, but also extends to the oligonucleotide. Residues in motif V both contact the oligonucleotide and line the interface between the first two domains. In particular, Thr-411 makes a hydrogen bond to the phosphate of dU3 of the oligonucleotide.

The current structure lacks any region corresponding to motif IV in Rep and PcrA helicases [H. S. Subramanya et al., Nature (1996), supra; S. Korolev et al., Cell, (1997), supra], members of the superfamily I

class of helicases. Previous sequence alignments that found similarities within motif IV between superfamily I and II helicases were done with rather weak criteria and may not have been significant. In the DNA helicases, motif IV is responsible for binding the adenine ring of ATP [H. S. Subramanya et al., Nature (1996), supra; S. Korolev et al., Cell, (1997), supra]. Mutation of a conserved arginine in this motif in UvrD increases the ATP Km by 37-fold [M. C. Hall et al., J. Biol. Chem., 272, pp. 18614-18620 (1997)].

In HCV helicase either another protein segment which is not in the current structure substitutes for motif IV or the adenosine ring binds elsewhere. Residues from the putative motif IV in HCV helicase include Ser-370 and Lys-371, which contact the DNA via a water-mediated hydrogen bond and a backbone interaction, respectively. Therefore sequences corresponding to motif IV in superfamily I and superfamily II helicases occupy different regions and appear to have different functions. We suggest that a new motif, designated IVa, be used to describe residues corresponding to the putative HCV helicase motif IV. In *E. coli* UvrD, motif IVa may correspond to the sequence RSNAQSRVL (residues 355-363).

Proposed domain closure and translocation

Conserved, basic residues from motif VI are positioned across the interdomain cleft from the expected location of the ATP γ phosphate in HCV helicase. A very similar situation is observed in the structures of the adenylate kinases, where basic residues lie across a cleft from the ATP binding site. Binding of ATP (or analogs) to these kinases leads to a conformational change in the enzyme, resulting in the burial of previously solvent-exposed phosphates [T. Bilderback et al., Biochemistry, 35, pp. 6100-6106 (1996)]. Mutation of these conserved basic residues in adenylate kinase results in an open structure with poor catalytic activity [G. E. Schulz, Faraday Discuss., 93, pp. 85-93 (1992)].

We propose that an analogous closure occurs between domains 1 and 2 of HCV helicase upon ATP binding. This closure could be driven by interaction of basic residues in motif VI with the ATP phosphates. Sequence and structural conservation of these basic residues in motif VI among superfamily I and II helicases suggests that domain closure upon ATP binding is a general feature of these enzymes.

Gln-460 and His-293, from motifs VI and II respectively, lie on opposite sides of the interdomain

cleft and possibly serve as gatekeepers, altering the equilibrium between the open and closed forms based on the binding of polynucleotide. Potential interaction of residues in these positions was predicted by the observation that helicases with the DExH motif II sequence usually contain a glutamine in motif VI, whereas those with a DEAD sequence contain a histidine [A. Gorbalenya et al., Curr. Opin. Struct. Biol. (1993), supra].

There is structural evidence that the linkage of the second domain in HCV helicase to the rest of the protein is flexible. In the HCV helicase structure reported by N. Yao et al., Nat. Struct. Biol. (1997), supra, the differences between the two molecules in the asymmetric unit can be attributed to a rotation of domain 2 by a few degrees. The relatively minor movement of domain 2 observed in their structures probably reflects changes in the local environment in the crystals. We propose a much more substantial conformational change would occur when the enzyme binds ATP or suitable analog. Our crystallographic results indicate that there are significant movements of this domain in different crystal forms. A conformational change could explain the observed two-stage kinetics of ATP binding to Rep where

rapid initial binding is followed by a much slower step, leading to tighter binding [K. J. Moore et al., Biochemistry, 33, pp. 14550-14564 (1994)]. Evidence for conformational changes have been observed for Rep and helicase II based on alterations in protease sensitivity upon nucleotide binding [K. Chao et al., J. Biol. Chem., 265, pp. 1067-1076 (1990)]. Binding of ATP to PcrA helicase has also been proposed to lead to a conformational change of the enzyme [H. S. Subramanya et al., Nature (1996), supra].

Large conformational changes in a DNA metabolizing enzyme are not unique, as they have been seen in the structures of mRNA capping enzyme in the presence of GTP [K. Hakansson et al., Cell, 89, pp. 545-553 (1997)]. In these structures the guanosine nucleotide is bound to the N-terminal domain with the phosphates located near the interface with the C-terminal domain. In the "open" conformation these domains are separated by a 10-13 Å cleft. Several residues which are highly conserved among mRNA capping enzymes are located in the C-terminal domain, including Arg-295 and Arg-298.

In the "closed" conformation, these residues are relocated by approximately 10 Å and are bound to the GTP β and γ phosphates. Closures of large interdomain

clefts have also been proposed in the structurally homologous ATP-dependent DNA ligases, of which one structure has been solved in the open conformation [H. S. Subramanya et al., Cell, 85, pp. 607-615 (1996)].

The second domain of HCV helicase also interacts with the single-stranded polynucleotide. One could envision that movement of this domain results in concomitant movement of the nucleic acid substrate relative to the protein. Interactions between residues in domain 2 such as Val-432 and Thr-448 and the bases at the 5' end of the single stranded nucleic acid binding site would lead to translocation of the polynucleotide in the 5' to 3' direction as domain 2 closes.

Trp-501 in domain 3 stacks with a base near the 3' end of the single stranded oligonucleotide and disrupts stacking with neighboring bases. Closure of the interdomain cleft would force several bases to slip past Trp-501. Hydrolysis of ATP would then result in opening of the cleft and release of ADP. The orientation of Trp-501 favors movement of the polynucleotide in only the 5' to 3' direction such that opening of the cleft results in net movement of domain 2 in a 3' to 5' direction. By this mechanism the translocation reaction of the helicase resembles a ratchet. A general ratchet-like mechanism

has been proposed for the RecB helicase based on conformational changes observed by protease mapping [R. J. Phillips et al., Mol. Gen. Genet., 254, pp. 319-329 (1997)].

Such a model suggests that a single ATP hydrolysis event can result in protein translocation of several bases along a polynucleotide. Studies with the UvrD DNA helicase have demonstrated that the enzyme is capable of translocating more than one base per reaction cycle [J. A. Ali et al., Science, 275, pp. 377-380 (1997)], although the number of ATP hydrolysis events per observed reaction cycle was unknown in this experiment. Our model is consistent with predictions that helicases need not actively unwind the double-stranded substrate, but can function by capturing the single-stranded regions which arise due to thermal fluctuations at the fork [Y. Z. Chen et al., J. Biomol. Struct. Dyn., 10, pp. 415-427 (1992)]. The translocation process proposed here would thus be considered an active process while the melting of double stranded structure at the fork would be passive.

The mechanism which we propose is substantially different from one described for the Rep helicase by Wong and Lohman [I. Wong et al., Science, 256, pp. 350-355 (1992)] and recently advanced in a paper describing the

3.0 and 3.2 Å structures of Rep bound to single-stranded DNA [S. Korolev et al., Cell (1997), supra]. As we previously noted there is overall structural similarity between domains 1 and 2 of HCV helicase and domains 1A and 2A of Rep. Important to our proposed mechanism, these two domains contain all the motifs conserved among DNA/RNA helicase sequences listed in Figure 9A. In HCV helicase, there is no structural equivalent of Rep domain 2B which has been proposed to have a critical role in the active rolling mechanism [S. Korolev et al., Cell (1997), supra].

EXAMPLE 5

Assays

A. Helicase assay

The standard 3'-tailed double-stranded RNA/DNA hybrid was prepared as described as follows. The long 98-nucleotide ("nt") RNA template was transcribed from a BsrBI-digested plasmid pSP65 (Promega, Madison, WI) in the presence of [α-^{32}P-GTP] (New England Nuclear, Boston, MA). The short 34-nt DNA release strand corresponds to a SP6 RNA transcript from a BamHI-digested pSP64 (Promega).

Standard helicase reactions (20 µl) were carried out as follows. HCV NS3 helicase (0.3 or 1 nM)

was added to a mixture of 25 mM morpholinepropanesulfonic acid (MOPS)-NaOH (pH 6.5), 1 mM ATP, 0.5 mM $MnCl_2$, 2 mM dithiothreitol (DTT), 0.1 mg of bovine serum albumin (BSA) per ml, 4 units of RNasin (Promega), and 5 nM of 3'-tailed double-stranded RNA/DNA hybrid substrate. Mixtures were incubated for 20 min at 37°C and stopped by the addition of 5 liters of 5X loading buffer [100 mM Tris-Cl (pH7.5), 20 mM EDTA, 50 % glycerol, 0.5 % SDS, 0.1 % NP-40, 0.1 % bromophenol blue, and 0.1 % xylene cyanole). The reactions were then electrophoresed on 10% PAGE with 0.5x TBE and 0.1 % SDS. Gels were dried and exposed using Fuji 1500 phosphorimager (Fuji, Stamford, CT). Helicase activity was determined by radioactivity of the double-stranded substrate and single-stranded template.

First, we characterized unwinding activity of the purified wild type NS3 helicase domain protein with regarding to the following parameters: protein concentration, incubation time course, incubation temperature, ATP concentration, pH, monovalent cation (Na^+), and divalent cation (Mn^{2+} and Mg^{2+}) (Fig. 11).

The helicase unwinding activity increased as the protein concentration or incubation time increased (Fig. 11A). At 0.1 nM of the NS3 helicase, the reaction

was almost linear with regard to the incubation time up to 30 min (Fig. 11A). Several NS3 helicase mutants purified by the same chromatograph method did not show any unwinding activity (see Table 1 and Example 6, below), indicating that the unwinding activity shown here is due to the purified HCV NS3 helicase, not containment proteins from *E. coli*.

Higher incubation temperature also led to more rapid unwinding of substrate (Fig. 11B), presumably due to lower energy required for break-down of hydrogen bonds between two strands at higher temperature. The unwinding activity of this NS3 helicase domain was optimal at pH 6.5, with a very narrow pH window of being active (Fig. 11C).

In addition, the unwinding reaction was very sensitive to the monovalent cation, such as Na^+ (Fig. 11D). Addition of 25 mM NaCl decreased the unwinding activity to about 15% of that in the absence of extra NaCl.

This helicase activity was absolutely dependent on the presence of ATP (Fig. 11E) or any other type of nucleotide triphosphate (NTP) (data not shown). The unwinding activity increased almost linearly as the concentration of ATP increased up to 1 mM (Fig. 11E).

However, at 5 mM of ATP, the unwinding activity is lower than that at 1 mM of ATP, probably due to inhibition of extra Na^+ brought in with the ATP.

The helicase activity also absolutely require the presence of divalent cations, such as Mn^{2+} or Mg^{2+} (Fig. 11F). However, if the concentration of divalent cation was higher than that of ATP, inhibition of the helicase activity was observed. At equal concentration of ATP and divalent cation (1 mM or 5 mM), Mn^{2+} showed higher unwinding activity than Mg^{2+} (Fig. 11F).

B. Single Stranded RNA binding assay

The binding of single stranded RNA ("ssRNA") to the HCV NS3 helicase was measured by a nitrocellulose filter binding assay. A 34-nt RNA transcript was generated from BamHI-digested pSP64 plasmid using SP6 RNA polymerase in the presence of [α-^{32}P-GTP]. Standard ssRNA binding reactions (40 µl) were carried out as follows.

HCV NS3 helicase domain protein (6.25 nM) was added to a mixture of 25 mM morpholinepropanesulfonic acid (MOPS)-NaOH (pH 7.0), 2 mM dithiothreitol (DTT), 0.1 mg of bovine serum albumin (BSA) per ml, 4 units of RNasin (Promega), and 5 nM of [^{32}P]-ssRNA substrate.

Mixtures were incubated for 15 min at 30°C and filtered through a pre-wet nitrocellulose membrane. The filter were washed twice with washing buffer [50 mM MOPS-NaOH (pH 7.0) and 1 mM EDTA], dried and quantified in scintillation counter.

Next, we determine several parameters in the ssRNA binding to the purified NS3 helicase using a filter binding assay (Fig. 12). The association of ^{32}P-labeled ssRNA to the NS3 helicase was very quick, usually close to completion within a couple minutes of incubation (data not shown). As shown in Fig. 12A, binding of ssRNA to the NS3 helicase is protein concentration-dependent. Under 8 nM of the NS3 helicase, the amount of ssRNA bound is a linear function of protein concentration (Fig. 12A, insert), and there is 0.445 molecule of ssRNA bound for every molecule of the NS3 helicase being present in the reaction. The maximal amount of ssRNA binding achieved in this reaction is about 94%. The K_d of the ssRNA-NS3 helicase complex is calculated to be 5.18 nM, at which the 50% of maximal binding of ssRNA to the NS3 helicase domain was observed.

We also measured the off rate constant of pre-formed ssRNA-NS3 helicase complex (Fig. 12B). In this case, ^{32}P-labeled ssRNA was incubated with the NS3

helicase protein was incubated together for 15 minutes to allow the formation of ^{32}P-ssRNA-NS3 helicase complex. Then 50-fold excess of ^{3}H-labeled ssRNA with the same sequence was added to the reaction so that any ^{32}P-labeled ssRNA dissociated from the complex with the NS3 helicase would have very little chance to re-associate with the NS3 protein again. The dissociation rate was determined to be 1.52×10^{-2} min-1.

We also examined effect of pH, monovalent (Na^{+}) and divalent (Mn^{2+}) cations on the ssRNA binding to the NS3 helicase. In contrast to the unwinding activity, ssRNA binding of the NS3 helicase was less sensitive to the pH change (Fig. 12C). The optimal binding was observed at pH 7.0, although ssRNA binding did not change significantly between pH 6.5 to 8.0. NaCl (Fig. 12D) and $MgCl_2$ (Fig. 12E) has an inhibitory effect on the ssRNA binding, although this inhibition curve as a function of salt concentration is not as sharp as that on unwinding activity.

C. ATPase assay

ATPase was measured by hydrolysis of ATP to ADP using a thin layer chromatography method [J.K. Tamura et. al., Virology, 193, pp.1-10 (1993)]. Standard ssRNA

binding reactions (10 µl) were carried out as follows. HCV NS3 helicase domain protein (2 nM) was added to a mixture of 50 mM morpholinepropanesulfonic acid (MOPS)-NaOH (pH 7.0), 0.1 mM ATP, 2.5 µCi of [α-^{32}P-ATP] (NEN), 0.5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 0.1 mg of bovine serum albumin (BSA) per ml, and in the presence or absence of 5 µM poly U (Uridine concentration) (Pharmacia). Mixtures were incubated for 30 min at 37°C and terminated by addition of EDTA to a final concentration of 20 mM. 0.5 µl of the reaction was spotted on a polyethyleneimine-cellulose plate, ATP and ADP were separated in 375 mM potassium phosphate (pH 3.5) and quantified by Fuji phosphorimager.

ATPase activity of the purified NS3 helicase was examined using this method. As shown in Fig. 13, this NS3 helicase domain has a basal ATPase activity in the absence of any polynucleotide, and the ATPase activity was stimulated up to 11-fold in the presence of poly(U). The order of ATPase stimulation by polynucleotides is poly(U) > poly(C) > poly(A) > poly(G) (data not shown).

EXAMPLE 6

Structure-Based Mutagenesis Study of RNA-Binding Residues of the NS3 Helicase

Mutagenesis experiments were performed to examine the roles of several residues predicted to be important in the NS3 helicase:oligonucleotide interaction based upon the crystal structure of that complex.

Ser231, Thr269, Ser370 and Thr411 formed direct water-mediated hydrogen bonds with the phosphate groups of the bound oligonucleotide. We replaced each one of these amino acids individuallv with alanine (see Example 1) and observed the effect of that mutation on various helicase activities. Alanine substitution at Ser231 or Ser370 had no observable effect on basal or polyU-stimulated ATPase activity, unwinding activity or ssRNA binding activity as compared to wild type helicase (see Table 1, below). Thus, it was concluded that those amino acids were not essential to define the oligonucleotide binding pocket in NS3 helicase.

In contrast, alanine substitution at Thr269 or Thr411 decreased the ssRNA binding to 20% of wild type level and completely abolished both polyU-stimulated ATPase activity and unwinding activity. Interestingly, basal ATPase activity was unaffected by either of these mutations.

The crystal structure also suggested that the side chain of Trp501 interacts with the bound

oligonucleotide. Substitution of this Trp with either Ala or Leu resulted in decreased ssRNA binding and abolished polyU-stimulated ATPase activity and unwinding activity, although basal ATPase activity was unaffected. In contrast a Trp501-to-Phe mutation did not affect basal ATPase, unwinding and ssRNA binding activities. This mutant was, however, less sensitive to polyU-stimulation of ATPase activity as compared to the wild type helicase. Surprisingly, the ATPase activity of this mutant when stimulated by other polynucleotides, such as polyC, was similar to that of the wild type.

Table 1. Mutational Study of Amino Acids in the RNA Binding Site of HCV NS3 Helicase

Amino Acid Mutation	Basal ATPase Activity (% of basal WT level)	Poly-U Stimulated ATPase Activity (% of basal WT level)	ssRNA Binding Activity (% of WT level)	ds RNA/DNA Unwinding Activity (% of WT level)
None (WT)	100	823	100	100
S231->A	260	709	121	99.8
T269->A	60	47	21	1
S370->A	104	694	124	109
T411->A	274	205	24	0.25
W501->F	99	197	100	112
W501->L	114	47	21	0.07
W501->A	101	49	40	0.36

Based upon these studies, it is apparent that Thr269, Thr411 and Trp501 are key residues for oligonucleotide binding. As indicated above, Thr269 and Trp501 make direct contacts with dU8. The minimal helicase amino acids which define the pocket in which dU8

lies are Val232, Thr254, Gly255, Thr269, Gly271, Lys272, Ala275, Trp501 and Tyr502. Thus, any compound which fits into a pocket comprising the structural coordinates ± a root mean square of 1.5 Å or less from the backbone atoms of these amino acids is a potential inhibitor of the NS3 helicase.

Additional amino acids that are located within 4 Å to 8 Å from the dU8 pocket are Pro230, Val256, Thr298, Ala497, Lys551, Gln552, Gly554, Glu555, Asn556 and Pro558. Thus, the combination of these amino acids with those above further define the dU8 pocket.

Based upon the crystal structure and these mutagenesis experiments, it is clear that Thr411 makes direct contact with dU4 and is a key residue in the U4 binding pocket. Other amino acids that are close enough to that U4 pocket to define its shape are His369, Ser370, Lys371, Tyr392, Arg393, Asp412, Ala413, Cys431, Val432, Gln434, Ile446, Thr448, Arg461, Glu493, Glu555, Asn556 and Phe557.

EXAMPLE 7

Structure-Based Mutagenesis Study of ATP-Binding Residues of the NS3 Helicase

Mutagenesis experiments were performed to examine the roles of several residues predicted to be

important in the NS3 helicase:ATP interaction based upon the crystal structure of that complex.

The mutations were achieved by the methods described in Example 1.

Table 2. Mutational Study of Amino Acids in the ATP Binding Site of HCV NS3 Helicase

Mutation	Basal ATPase (% of basal wt level)	Poly U-stimulated ATPase (% of basal wt level)	ssRNA binding (% of wt level)	dsRNA /DNA unwinding (% of wt level)
wild-type	100	581	100	100
Q460->A	23	32	97	3
R461->A	140	193	57	2
R462->A	247	337	99	81
R464->A	33	21	105	<0.01
R467->A	7	14	116	<0.05

In our model, R464 and R467 were predicted to bind to the γ- and α- phosphate groups of NTP, respectively. This is in contrast to what has previously been reported in the art, wherein these residues were predicted to be involved in RNA binding [T. Yao et al., Nat. Struct. Biol., 4, pp. 463-467 (1997); C. Hyun-Soo et al., J. Biol. Chem., 273, pp. 15045-15052 (1998)].

Two individual mutations, R464->A and R467->A, showed very low basal and polyU-stimulated ATPase activities. Although they had normal RNA binding ability, which suggested that the mutated protein has a proper fold, helicase unwinding activity was almost non-existent in these two mutant proteins, presumably due to the loss of NTPase activity. These results indicated

that these two Arg residues are critical for NTPase activity.

The Q460->A mutation had a similar effect as two above-mentioned Arg-to-Ala mutations. This Gln was predicted to interact with and maintain the proper conformation of the imidazole ring of His-293 of the DECH motif.

The R461->A mutation led to lower RNA binding and less polyU stimulation of ATPase activity, which resulted in a very low helicase unwinding activity.

The R462->A mutation had no major effect on any of these four activities as predicted.

While we have hereinbefore presented a number of embodiments of this invention, it is apparent that my basic construction can be altered to provide other embodiments which utilize the methods of this invention. Therefore, it will be appreciated that the scope of this invention is to be defined by the claims appended hereto rather than the specific embodiments which have been presented hereinbefore by way of example.